

Correction to the 2005 paper: "Digit Selection for SRT Division and Square Root"

Peter Kornerup
Dept. of Mathematics and Computer Scienc
University of Southern Denmark, Odense, Denmark
E-mail: kornerup@imada.sdu.dk

Abstract

It has been pointed out by counterexamples in a 2013 paper in the IEEE Transactions on Computers [1], that there is an error in the previously ibid. in 2005 published paper [2] on the construction of valid digit selection tables for SRT type division and square root algorithms. The error has been corrected, and new results found on selection constants for maximally redundant digit sets.

Index Terms

Digit selection, SRT, division, square root

I. INTRODUCTION

In a recent paper [1], David M. Russinoff expressed criticism on the determination of digit selection parameters in SRT division algorithms, presented in the paper [2] by this author. An error in the selection was pointed out by counterexamples.

The SRT algorithms for division and square root are based on selecting the digits of the result by a table look-up or equivalent, using a few leading bits of the divisor (or root approximation) and of the partial remainder.

To determine the minimal number of bits necessary for a valid table to exist, traditionally searches were performed to assure that the next quotient digit can be chosen as valid for all points (remainder, divisor) in a set defined by the truncated remainder and divisor, i.e., a specific uncertainty rectangle.

It was the purpose of [2] as an alternative to present a more analytical approach to determine these parameters, based directly on the radix and digit set of the quotient/root representation. Below is a brief account of the core of this approach, but with some additional considerations on these parameters, followed by an analysis of the error in [2], including a correction for its parameter determination and simplifications when the digit set is maximally redundant. Finally some conclusions are presented.

II. PARAMETER DETERMINATION IN [2]

Let β be the radix (assumed to be a power of 2) and $\{-a, \dots, a\}$ be the digit set of the quotient, where $\beta/2 \leq a \leq \beta - 1$ and $\rho = \frac{a}{\beta-1}$ is the redundancy index. Let t and u to be determined be respectively the number of leading fractional digits of the partial remainder and of the divisor y , assumed normalized $1/2 \leq y < 1$. The digit selection is to be based on a table look-up (or equivalent) essentially indexed by t and u , the table size then being exponential in $u + t$. Hence we seek t and u such that $u + t$ is minimal, normally obtained by minimizing u , which in [3] by synthesis studies has been confirmed also to generally minimize the delay and area.

An analysis on the positioning of the "uncertainty rectangles" leads to the following condition (Equation (12) from [2]) for a valid digit selection table to exist:

$$\begin{aligned} & \left\lceil 2^{t-u}(d - \rho)k + 2^{t-u}(d - \rho) + 1 \right\rceil \\ & \leq \lfloor 2^{t-u}(d - \rho)k + 2^{t-u}(2\rho - 1)k \rfloor, \end{aligned} \quad (1)$$

which has to be satisfied for all k , $2^{u-1} \leq k < 2^u$ and digits $d > 0$ (which can be assumed by symmetry). Note that the inner leftmost terms in the two sides of the inequality are identical, and thus the condition essentially depends on the rightmost terms. Also note that the significance of the terms is increased by maximizing the difference $t - u$.

It is then seen that (1) is satisfied if the following:

$$2^{t-u}((2\rho - 1)k - (d - \rho)) \geq 2,$$

holds for the minimal value $k = 2^{u-1}$ and the maximal value $d = a$, and thus also for all $d < a$ and $k > 2^{u-1}$. This translates into the condition:

$$2^{-t} \leq \left(\left(\rho - \frac{1}{2} \right) - (a - \rho)2^{-u} \right) / 2 \quad (2)$$

which may be used to find values of u and t . However, there is a chance that the inequality (1) can be satisfied even if the weaker condition:

$$2^{-t} \leq \left(\left(\rho - \frac{1}{2} \right) - (a - \rho)2^{-u} \right) \quad (3)$$

similarly is satisfied.

For any of these conditions to hold, it is obviously necessary that u is chosen such that:

$$2^{-u} < \frac{\rho - \frac{1}{2}}{a - \rho}, \quad (4)$$

provided that $a > \rho$, or $\beta > 2$, since $\beta = 2$ is the only case where $\rho = a (= 1)$, a case which can be handled separately.

Given any value of u satisfying (4), a possible value $t = t_0$ can then be determined say from (3) as:

$$t_0 = \left\lceil \left(-\log_2 \left(\rho - \frac{1}{2} - (a - \rho)2^{-u} \right) \right) \right\rceil, \quad (5)$$

however, it may be necessary to apply the stronger condition (2), in which case $t = t_0 + 1$. To decide between these two situations the difference between the righthand and lefthand expressions in (1) may be checked for given specific values of u and t_0 .

Note that u by (4) can be chosen arbitrarily large, and that t_0 by (5) decreases when $u \rightarrow \infty$, e.g., for maximally redundant digit sets ($\rho = 1$), $t_0 \rightarrow 2$. Hence the factor 2^{t-u} in (1) can be made arbitrarily small. However, we want $u + t$ to be small to minimize the table.

III. THE ERROR AND ITS CORRECTION

Russinoff in [1] points out by counterexamples for large values of the radix and u , that the test in [2], fails in some cases to correctly identify whether to use $t = t_0$ or $t = t_0 + 1$.

The test is based on checking whether the difference $\Delta(t, u, d, \rho, k)$ between the expressions in (1) is non-negative for $2^{u-1} \leq k < 2^u$, together with the (false!) observation that it is sufficient to perform it only for the digit value $d = a$, to assure that the inequality holds for all values of $d > 0$. If tested for all values of d it is equivalent to a check on the correct positioning of the “uncertainty rectangles” between some slanted lines. The counterexamples were found for radix 16 and 32 with $u = 9$, respectively $u = 11$ and $t_0 = 2$, which erroneously had accept for $d = a$ but failed for $d = a - 1$. Note that in both cases the value of 2^{t-u} is very small.

Let $\delta_{kd} = 2^{t_0-u} ((2\rho - 1)k - (d - \rho)) - 1$ be the difference between the internal expressions in (1). The test according to Theorem 3 of [2] fails for certain extreme combinations of u and t_0 ($u \gg t_0$), since the determination of t from (5) does not assure that $\delta_{kd} \geq 1$. When $\delta_{kd} < 1$ the above observation on the sufficiency of the test for $d = a$ does not hold. Note that for $u \gg t_0$, δ_{kd} grows only slowly with k . In the two counterexamples with $t_0 = 2$ it is found that:

$$\begin{aligned} \beta = 16 \quad u = 9 \quad d = a = 15 \quad k = 2^{u-1} \quad \delta_{kd} &= 0.890625 \\ \beta = 32 \quad u = 11 \quad d = a = 31 \quad k = 2^{u-1} \quad \delta_{kd} &= 0.94140625. \end{aligned}$$

If $0 < \delta_{kd} < 1$ and the two internal expressions in (1) considered as an interval happens to include an integer value x for some value of k , then $\Delta(t, u, d, \rho, k) = 0$. But for each increment of k , the left endpoint of the interval will be shifted to the right by an amount $2^{t_0-u}(d - \rho)$ and the width increased by $2^{t_0-u}(2\rho - 1)$. Eventually, if still $\delta_{kd} < 1$, it may fall in the open interval between x and $x + 1$, then $\Delta(t, u, d, \rho, k) = -1$ and the test fails. If this happens for $d = a$ then Theorem 3 in [2] specifies that $t = t_0 + 1$ should be used.

If a smaller value of t is wanted, a value larger than the minimal values of u may be used, not necessarily minimizing $u + t$. From the stronger condition (2), for any value of $t \geq \lceil -\log_2(\rho - \frac{1}{2}) \rceil + 1$, we find that

$$2^{-u'} \leq \frac{\rho - \frac{1}{2} - 2^{1-t}}{a - \rho}, \quad (6)$$

implies $\delta_{kd} \geq 1$ for $k = 2^{u'-1}$, and hence $\Delta(t, u, d, \rho, k) \geq 0$ for all $k > 2^{u'-1}$ and $d \leq a$. From (6) we may determine a u' which may be greater than the minimal u chosen by (4).

Thus in the above counterexamples, $t = 3$ is the minimal value possible for $\beta = 16$ and $\beta = 32$. Then $u' = 6$ respectively $u' = 7$ are the minimal values which could be used, and choosing these values we find:

$$\begin{aligned} \beta = 16 \quad u' = 6 \quad d = a = 15 \quad k = 2^{u'-1} \quad \delta_{kd} &= 1.25 \\ \beta = 32 \quad u' = 7 \quad d = a = 31 \quad k = 2^{u'-1} \quad \delta_{kd} &= 1.125, \end{aligned}$$

where the test accepts.

IV. THE CORRECTION TO [2]

Theorem 3 in [2] is based on the chance that the weaker condition (3) is sometimes sufficient to satisfy condition (1), and thus the smaller value $t = t_0$ can be used. But exhaustive searches for $u = u_0$ being the minimal solution to (4), has shown that this turns out to be the case in only very few cases. However, it has turned out that no searches are necessary in the case when the digit set is maximally redundant, hence we will deal with this case separately below.

It turns out that the test for all d whether $t_0 = \hat{t}$ on $\Delta(\hat{t}, u_0, d, \rho, k)$ for a restricted set of radices is only satisfied for $\beta = 4, a = 2$, for $\beta = 16, a = 10$, for $\beta = 32, a = 25$, for $\beta = 64, a = 38, 42, 44, 46, 51$, and for $\beta = 128, a = 81, 89, 94, 105$. In a few other cases the tests falsely indicated accept. The test on $\Delta(\hat{t}, u_0, d, \rho, k)$ generally fails when $2^{\hat{t}-u_0}$ is very small, but in no systematic way. A Maple program for the general determination of valid parameters is available.

A corrected and reorganized version of Theorem 3 of [2], now identifying further valid parameter pairs (u, t) , is then:

Theorem 1: (SRT digit selection constants)

For p -bit radix β SRT division for $\beta = 2^p, p = 2, \dots, 7$ with digit set $D = \{-a, \dots, a\}$, $\beta/2 \leq a < \beta - 1$, and $\rho = \frac{a}{\beta-1}$, the selection constants $\widehat{S}_d(\hat{y}) = s_{d,k}2^{-t}$ can be determined for $1 \leq d \leq a$ and $\hat{y} = k \cdot \text{ulp}(\hat{y})$ as

$$s_{d,k} = \left\lceil 2^{t-u}(d - \rho)(k + 1) \right\rceil$$

for $k = 2^{u-1}, \dots, 2^u - 1$, using truncation parameters t, u defined by $\text{ulp}(\widehat{S}_d(\hat{y})) = \text{ulp}(\widehat{\beta r}_i) = 2^{-t}$ and $\text{ulp}(\hat{y}) = 2^{-u}$, where u has to satisfy

$$2^{-u} < \frac{\rho - \frac{1}{2}}{a - \rho}. \quad (7)$$

If $u = u_{\min}$ is the minimal value satisfying (7), then let t' be the smallest value of t satisfying

$$t > 1 - \log_2(\rho - \frac{1}{2}), \quad (8)$$

and define $u = u_{\max}$ as the smallest value of u satisfying

$$2^{-u} \leq \frac{\rho - \frac{1}{2} - 2^{1-t'}}{a - \rho}. \quad (9)$$

For any value of u , $u_{\min} \leq u \leq u_{\max}$ define \hat{t} as the smallest value of t satisfying

$$2^{-t} \leq (\rho - \frac{1}{2}) - (a - \rho)2^{-u}, \quad (10)$$

and define from (1)

$$\Delta(t, u, d, \rho, k) = \left\lfloor 2^{t-u}(d + \rho - 1)k \right\rfloor - \left\lfloor 2^{t-u}(d - \rho)(k + 1) \right\rfloor + 1.$$

Also define the following two checks:

$$\text{simple} = \exists k \in \{2^{u-1} \dots 2^u - 1\} : \Delta(\hat{t}, u, a, \rho, k) < 0$$

and

$$\begin{aligned} \text{rest} = \exists k \in \{2^{u-1} \dots 2^u - 1\} \text{ and } \exists d \in \{0 \dots a - 1\} \\ : \Delta(\hat{t}, u, d, \rho, k) < 0 \end{aligned}$$

Then

$$t = \begin{cases} \hat{t} + 1 & \text{if } \text{simple}, \\ \hat{t} + 1 & \text{if } \neg \text{simple} \wedge \text{rest}, \\ \hat{t} & \text{otherwise,} \end{cases}$$

then (u, t) provides a set of parameters defining a valid digit selection table.

Proof: The expression for $s_{d,k}$ is from (1), and the condition (7) on u is necessary, from which the minimal value u_{min} is derived. Comparing (7) with (9) it is seen that $u_{max} \geq u_{min}$.

The only situations where $t = \hat{t}$ can be verified, given β and a , are when $\Delta(\hat{t}, u, d, \rho, k) \geq 0$ for all $d \in \{1, \dots, a\}$ and $k \in \{2^{u-1}, \dots, 2^u - 1\}$, yielding the combinations listed. The split cases when \hat{t} must be increased covers situations where the test fails for some value of d and k , and the strong condition (2) must be applied. ■

It is only necessary to check if $\Delta(\hat{t}, u, d, \rho, k) \geq 0$ for $d \in \{1 \dots, a-1\}$ and all k when $\Delta(\hat{t}, u, d, \rho, k) \geq 0$ for all k . This is where the original theorem failed by only testing the latter. But observe that if the simple test turns out false, no further testing is necessary. As mentioned above there are only very few situations where $t = \hat{t}$.

Also note that no solutions are possible for $u < u_{min}$, and if (u, t) is a valid pair, then $(u + s, t)$ and $(u, t + s)$ for any $s > 0$ are also, but obviously not as good.

Example 1 With the minimally redundant digit set for $\beta = 16$, $a = 8$, $u_{min} = 8$ and $u_{max} = 12$. For the possible choices of u we find:

u	t	$u + t$
8	9	17
9	7	16
10	7	17
11	7	18
12	6	18

where $(u, t) = (9, 7)$ yields the minimal value of $u + t$. ■

Theorem 2: (SRT for maximally redundant digit sets)

For $\beta = 2^p$, $p > 2$, with the maximally redundant digit set $D = \{-\beta + 1, \dots, 0, \dots, \beta - 1\}$ there are two sets of parameters (u, t) defining valid digit selection tables:

u	t	$u + t$
$u_{min} = p + 1$	p	$2p + 1$
$u_{max} = p + 2$	3	$p + 5$

For $p = 2$, $u = u_{min} = u_{max}$ there is only one set:

u	t	$u + t$
3	2	5.

Proof: With $a = \beta - 1$, $\rho = 1$, for $\beta = 2^p$ it follows that $u_{min} = \lceil \log_2(2^p - 2) + 1 \rceil = p + 1$. Then $t' = 3$, from which $2^{u_{max}} \geq 2^{p+2} - 8$, implying $u_{max} = p + 2$ when $p > 2$, but for $p = 2$ $u_{max} = 3$, which is identical to u_{min} . From $2^{-t} \leq (\rho - \frac{1}{2}) - (a - \rho)2^{-u}$ for $u = u_{min} = p + 1$ the minimal t is $\hat{t}_{min} = p$, and for $u = u_{max} = 3$, $\hat{t}_{max} = 2$.

For $(u, t) = (u_{min}, \hat{t}_{min}) = (p + 1, p)$:

$$\begin{aligned} \Delta(\hat{t}_{min}, u_{min}, d, 1, k) &= \lfloor 2^{-1}dk \rfloor - \lfloor 2^{-1}dk + 2^{-1}(d - k - 1) + 1 \rfloor \\ &\geq \lfloor 2^{-1}d'k' \rfloor - \lfloor 2^{-1}d'k + 2^{-1}((2^p - 1) - (2^p + 1) - 1) + 1 \rfloor \\ &= \lfloor 2^{-1}d'k' \rfloor - \left\lceil 2^{-1}d'k' - \frac{1}{2} \right\rceil = 0, \end{aligned}$$

when substituting d by its maximal value $d' = 2^p - 1$ and k by its (almost) minimal value $k' = 2^{u_{min}-1} + 1 = 2^p + 1$, using that with these extreme values $d'k'$ is odd. Substituting with $k'' = 2^{u_{min}-1} = 2^p$ then $d'k''$ is even and the lower bound is $\lfloor 2^{-1}d'k'' \rfloor - \lfloor 2^{-1}d'k'' \rfloor = 0$.

Hence $(u, t) = (u_{min}, \hat{t}_{min}) = (p + 1, p)$ provides a correct table, which also covers the case for $p = 2$ with $(u, t) = (3, 2)$.

For $(u, t) = (u_{max}, \hat{t}_{max}) = (p+2, 2)$ we will now show that there is a value $k = 2^{u_{max}-1} + 2 = 2^{p+1} + 2$ such that $\Delta(t, u, a, \rho, k) < 0$ for $d = a = 2^p - 1$. Let $K = 2^{\hat{t}_{max}-u_{max}} a k = 2(2^p - 2^{-p})$ then

$$\begin{aligned}\Delta(\hat{t}_{max}, u_{max}, a, 1, k) &= \lfloor K \rfloor - \lceil K + 2^{-p}(a - k - 1) + 1 \rceil \\ &= \lfloor K \rfloor - \lceil K + 2^{-p}(2^p - 1 - 2(1 + 2^p) + 1) + 1 \rceil \\ &= \lfloor K \rfloor - \lceil K - 2^{-p} \rceil = (2^{p+1} - 1) - 2^{p+1} = -1.\end{aligned}$$

Thus \hat{t}_{max} must be incremented and $(u, t) = (p+2, 3)$ provides a correct table. ■

Example 2 For the maximally redundant digit set with $\beta = 16$, $a = 15$, the minimal value of u is $u_{min} = 5$, for which $t' = \hat{t} = 4$ is determined. However, a smaller value of t , $t'' = 3$ is also possible, for which $u_{max} = 6$. Note that $u + t$ is the same for the two combinations, hence they require the same table sizes, but when t is smaller, fewer bits of the redundant partial remainder need to be converted. ■

V. CONCLUSIONS

It is likely that the error had not been noticed because it was implicitly assumed that minimal tables are wanted, as obtained by choosing minimal or almost minimal values of u , and thus for small values of $u + t$ and small values of $u - t$. Fortunately, the exposure of the error has prompted a further analysis of the problem of determining additional value pairs (u, t) , providing valid digit selection tables. The result on truncation parameters u, t for the more general case has been significantly strengthened. A new theorem is presented, simplifying the parameter determination in the important case when the digit set is maximally redundant, eliminating all searching.

As pointed out by Russinoff, the error has gone unnoticed in the review process, and subsequently by other referencing the paper. However it is standard scientific knowledge, that any research result has to prove its correctness through the “time test”, i.e. that it can stand uncontested through time. It is appreciated that his objections identified the problem, and made it possible in this case to provide a correction of the presented results.

He further contests the use of “informal quasi-mathematical arguments” as opposed to a “formal machine-checked proof”, such as he has applied to his proofs in [1], employing an ACL2 proof script which consists of more than 800 lemmas, an impressive effort.

His approach is the same as in publications before [2] to determine parameters for providing valid digit selection tables: proving the validity of some pair (u, t) by checking all table entries, but limited to maximally redundant digit sets. The attempt in [2] was to determine these parameters directly, based on the radix and any corresponding valid digit set, and in this revision has been significantly strengthened.

REFERENCES

- [1] D. Russinoff, “Computation and Formal Verification of SRT Quotient and Square Root Digit Selection Tables,” *IEEE Transactions on Computers*, vol. 62, no. 5, pp. 900–913, May 2013.
- [2] P. Kornerup, “Digit Selection for SRT Division and Square Root,” *IEEE Transactions on Computers*, vol. 54, no. 3, pp. 294–303, March 2005.
- [3] S. Oberman and M. Flynn, “Minimizing the Complexity of SRT Tables,” *IEEE Transactions on VLSI systems*, vol. 6, no. 1, pp. 141–149, March 1998.